Genome Glossary

A

Adenine (A) A nitrogenous base, one member of the base pair AT (adenine and thymine).

Allele Alternative form of a genetic locus; a single allele for each locus is inherited separately from each parent (e.g.,at a locus for eye color the allele might result in blue or brown eyes).

Amino acid Any of a class of 20 molecules that are combined to form proteins in living things. The sequence of amino acids in a protein and hence protein function are determined by the genetic code.

Amplification An increase in the number of copies of a specific DNA fragment; can be in vivo or in vitro. See cloning, polymerase chain reaction.

Arrayed library Individual primary recombinant clones (hosted in phage, cosmid, YAC, or other vector) that are placed in two-dimensional arrays in microtiter dishes. Each primary clone can be identified by the identity of the plate and the clone location (row and column) on that plate. Arrayed libraries of clones can be used for many applications, including screening for a specific gene or genomic region of interest as well as for physical mapping. Information gathered on individual clones from various genetic linkage and physical map analyses is entered into a relational database and used to construct physical and genetic linkage maps simultaneously; clone identifiers serve to interrelate the multilevel maps. Compare library, genomic library.

Autosome A nitrogenous base, one member of the base pairAT (adenine and thymine).

B

BAC See bacterial artificial chromosome.

Bacterial artificial chromosome (BAC) A vector used to clone DNA fragments (100- to 300-kb insert size; average, 150 kb) in *Escherichia coli* cells. Based on naturally occurring F-factor plasmid found in the bacterium *E. coli*. Compare cloning vector.

Bacteriophage See phage.

Base pair (bp) Two nitrogenous bases (adenine and thymine or guanine and cytosine) held together by weak bonds. Two strands of DNA are held together in the shape of a double helix by the bonds between base pairs.

Base sequence The order of nucleotide bases in a DNA molecule.

Base sequence analysis A method, sometimes automated, for determining the base sequence.

bp See base pair.

C

cDNA See complementary DNA.

Centimorgan (cM) A unit of measure of recombination frequency. One centimorgan is equal to a 1% chance that a marker at one genetic locus will be separated from a marker at a second locus due to crossing over in a single generation. In human beings, 1 centimorgan is equivalent, on average, to 1 million base pairs.

Chromosome The self-replicating genetic structure of cells containing the cellular DNA that bears in its nucleotide sequence the linear array of genes. In prokaryotes, chromosomal DNA is circular, and the entire genome is carried on one chromosome. Eukaryotic genomes consist of a number of chromosomes whose DNA is associated with different kinds of proteins.

Clone bank See genomic library.

Clone A group of cells derived from a single ancestor.

Cloning The process of asexually producing a group of cells (clones), all genetically identical, from a single ancestor. In recombinant DNA technology, the use of DNA manipulation procedures to produce multiple copies of a single gene or segment of DNA is referred to as cloning DNA.

Cloning vector DNA molecule originating from a virus, a plasmid, or the cell of a higher organism, into which another DNA fragment of appropriate size can be integrated without loss of the vector's capacity for self-replication; vectors introduce foreign DNA into host cells, where it can be reproduced in large quantities. Examples are plasmids, cosmids, and yeast artificial chromosomes; vectors are often recombinant molecules containing DNA sequences from several sources.

Code See genetic code.

Codon See genetic code.

Complementary DNA (cDNA) DNA that is synthesized from a messenger RNA template; the single-stranded form is often used as a probe in physical mapping.

Complementary sequence Nucleic acid base sequence that can form a double-stranded structure by matching base pairs with another sequence; the complementary sequence to GTAC is CATG.

Conserved sequence A base sequence in a DNA molecule (or an amino acid sequence in a protein) that has remained essentially unchanged throughout evolution.

Contig A group of overlapping clones spanning a region of a genome.

Contig map A map depicting the relative order of a linked library of small overlapping clones representing a complete chromosomal segment.

Cosmid Artificially constructed cloning vector containing the cos gene of phage lambda. Cosmids can be packaged in lambda phage particles for infection into E. coli; this permits cloning of larger DNA fragments (up to 45kb) than can be introduced into bacterial hosts in plasmid vectors.

Crossing over The breaking during meiosis of one maternal and one paternal chromosome, the exchange of corresponding sections of DNA, and the rejoining of the chromosomes. This process can result in an exchange of alleles between chromosomes. Compare recombination.

Cytosine (C) A nitrogenous base, one member of the base pair GC (guanine and cytosine).

D

Deoxyribonucleotide See nucleotide.

Diploid A full set of genetic material, consisting of paired chromosomes, one chromosome from each parental set. Most animal cells except gametes have a diploid set of chromosomes. The diploid human genome has 46 chromosomes. Compare haploid.

DNA (deoxyribonucleic acid) The molecule that encodes genetic information. DNA is adouble-stranded molecule held together by weak bonds between base pairs of nucleotides. The four nucleotides in DNA contain the bases adenine (A),guanine (G), cytosine (C), and thymine (T). In nature, base pairs form only between A and T and between G and C; thus the base sequence of each single strand can be deduced from that of its partner.

DNA probe See probe.

DNA replication The use of existing DNA as a template for the synthesis of new DNA strands. In humans and other eukaryotes, replication occurs in the cell nucleus.

DNA sequence The relative order of base pairs, whether in a fragment of DNA, a gene, a chromosome, or an entire genome. See base sequence analysis.

Double helix The shape that two linear strands of DNA assume when bonded together.

E

E. coli Common bacterium that has been studied intensively by geneticists because of its small genome size, normal lack of pathogenicity, and ease of growth in the laboratory.

Electrophoresis A method of separating large molecules (such as DNA fragments or proteins) from a mixture of similar molecules. An electric current is passed through a medium containing the mixture, and each kind of molecule travels through the medium at a different rate, depending on its electrical charge and size. Separation is based on these differences. Agarose and acrylamide gels are the media commonly used for electrophoresis of proteins and nucleic acids.

Endonuclease An enzyme that cleaves its nucleic acid substrate at internal sites in the nucleotide sequence.

Enzyme A protein that acts as a catalyst, speeding the rate at which a biochemical reaction proceeds but not altering the direction or nature of the reaction.

EST Expressed sequence tag. See sequence tagged site.

Eukaryote Cell or organism with membrane-bound, structurally discrete nucleus and other well-developed subcellular compartments. Eukaryotes include all organisms except viruses, bacteria, and blue-green algae. Compare prokaryote. See chromosome.

Evolutionarily conserved See conserved sequence.

Exon The protein-coding DNA sequence of a gene. Compare intron.

Exonuclease An enzyme that cleaves nucleotides sequentially from free ends of a linear nucleic acid substrate.

Expressed gene See gene expression.

F

FISH (fluorescence *in situ* hybridization) A physical mapping approach that uses fluorescein tags to detect hybridization of probes with metaphase chromosomes and with the lesscondensed somatic interphase chromatin.

G

Gene The fundamental physical and functional unit of heredity. A gene is an ordered sequence of nucleotides located in a particular position on a particular chromosome that encodes a specific functional product (i.e., a protein or RNA molecule). See gene expression.

Gene expression The process by which a gene's coded information is converted into the structures present and operating in the cell. Expressed genes include those that are transcribed into mRNA and then translated into protein and those that are transcribed into RNA but not translated into protein (e.g., transfer and ribosomal RNAs).

Gene family Group of closely related genes that make similar products.

Gene library See genomic library.

Gene mapping Determination of the relative positions of genes on a DNA molecule (chromosome or plasmid) and of the distance, in linkage units or physical units, between them.

Gene product The biochemical material, either RNA or protein, resulting from expression of a gene. The amount of gene product is used to measure how active a gene is; abnormal amounts can be correlated with disease-causing alleles.

Genetic code The sequence of nucleotides, coded in triplets (codons) along the mRNA, that determines the sequence of amino acids in protein synthesis. The DNA sequence of a gene can be used to predict the mRNA sequence, and the genetic code can in turn be used to predict the amino acid sequence.

Genetic engineering technology See recombinant DNA technology.

Genetic map (linkage map) A map of relative positions of genes or other chromosome markers, determined on the basis of how often they are inherited together. See Physical map.

Genetic material See genome.

Genetics The study of the patterns of inheritance of specific traits.

Genome All the genetic material in the chromosomes of a particular organism; its size is generally given as its total number of base pairs.

Genome project Research and technology development effort aimed at mapping and sequencing some or all of the genome of human beings and other organisms.

Genomic library A collection of clones made from a set of randomly generated overlapping DNA fragments representing the entire genome of an organism. Compare library, arrayed library.

Guanine (G) A nitrogenous base, one member of the base pair GC (guanine and cytosine).

H

Haploid A single set of chromosomes (half the full set of genetic material), present in the egg and sperm cells of animals and in the egg and pollen cells of plants. Human beings have 23 chromosomes in their reproductive cells. Compare diploid.

Homology Similarity in DNA or protein sequences between individuals of the same species or among different species.

Human gene therapy Insertion of normal DNA directly into cells to correct a genetic defect.

Human Genome Initiative Collective name for several projects begun in 1986 by DOE to (1) create an ordered set of DNA segments from known chromosomal locations, (2) develop new computational methods for analyzing genetic map and DNA sequence data, and (3) develop new techniques and instruments for detecting and analyzing DNA. This DOE initiative is now known as the Human Genome Program. The national effort, led by DOE and NIH, is known as the Human Genome Project.

Hybridization The process of joining two complementary strands of DNA or one each of DNA and RNA to form a double-stranded molecule.

I

Informatics The study of the application of computer and statistical techniques to the management of information. In genome projects, informatics includes the development of methods to search databases quickly, to analyze DNA sequence information, and to predict protein sequence and structure from DNA sequence data.

In situ hybridization Use of a DNA or RNA probe to detect the presence of the complementary DNA sequence in cloned bacterial or cultured eukaryotic cells.

JGI
JOINT GENOME INSTITUTE